

פוסט 9: שיטות להערכה חלופית (לסקרי ההוראה) של ההוראה האקדמית – חלק ג: האם ניתן להשתמש באילו מהשיטות כתחליף לסקרי ההוראה?

בפוסטים 7 ו-8 הכרנו כ-15 שיטות חלופיות להערכת ההוראה האקדמית וראינו שלאף אחת מהן אין היתכנות - בערכי עלות/תועלת - ליישום בהיקף נרחב של מאות ואלפי קורסים בכל מועד הערכה. ואולם בנוסף לשיקולי עלות/תועלת, יש סיבה נוספת למניעת יישום כזה שהיא עקרונית לגבי שימוש בכלי מדידה, והיא בעיית המהימנות והתקפות של השיטות החלופיות.

III. בעיות בתקפות או מהימנות של השיטות החלופיות

ככל הידוע לי, אין אפילו אחת מבין השיטות החלופיות שנמצאה באופן מבוסס מדעית כמקיימת ביחד את שתי הדרישות ההכרחיות לכל כלי מדידה – מהימנות וגם תקפות! לפי לינס (Linse, 2017), כמעט אין אף אחת מהן שנבדקה באופן יסודי לגבי מהימנותה ותקפותה. ייתכן מאוד שחוסר הבדיקה נגרם משום שהחוקרים הבינו מראש שעצם המהות של דרך היישום של השיטה שולל את ההיתכנות של מהימנות או תקפות. להלן הסיבות המשוערות לגבי ההימנות מהבדיקה:

לגבי הביסוס של מהימנות. היות וכמתואר בפוסט 8 היישום של כל אחת מהשיטות הוא יקר מאוד מבחינת כסף, זמן, וכוח אדם, הרי שיקר כפליים - לשם ביסוס של מהימנות - לחזור על ההליך בו זמנית לגבי מספר מתאים של קורסים עם צוות אחר, כדי להראות שהתוצאות בשתי הפעמים שוות או קרובות זו לזו.

לשם ביסוס של תקפות. יש להראות שהשיטה אכן מודדת את מה שאנו מצפים מהוראה טובה, ושאינן הטיות בתוצאות ההערכה עקב סיבות שאינן רלוונטיות להוראה ולמידה. ואכן לגבי הנקודה השנייה, ההערכה בכל אחת מבין השיטות החלופיות ברשימה מפוסט 7, אשר נעשת במסגרת של אחד-על-אחד על ידי אנשי אקדמיה בעלי עניין (כמו המרצה עצמו בדירוג עצמי, עמיתים, ממונים ואדמיניסטרטורים, פורטפוליו), היא בהכרח סובייקטיבית ולכן עם הסתברות גבוהה להיות מוטת. בנוסף, הערכות על בסיס של תצפיות מבוססות במרבית המקרים (שוב, משיקולי עלות) על תצפית יחידה שאינה משקפת את כלל השיעורים במהלך הסמסטר, מה עוד שהמרצה יודע מראש לפני השיעור שתתקיים בשיעור זה תצפית ועלול להתנהג בשיעור הנצפה באופן שונה מהרגיל ובלתי טיפוסי.

IV. השוואת תוצאות סקרי ההוראה לתוצאות של הערכה בשיטות חלופיות

דרך נוספת לביסוס תקפות של הערכת מרצים בשיטה חלופית היא להראות שהערכה זו נותנת תוצאות דומות לאלו של הערכת מרצים בשיטה אחרת שנמצאה להיות תקפה. שיטת ההערכה היחידה שתוקפה עד כה היא דירוג הסטודנטים את המרצים (בתנאי שכל ההליך של סקר ההוראה עשוי בצורה נכונה וראויה). יש ממצאים המראים קורלציות בינוניות-עד-גבוהות של סקרי ההוראה עם שיטות חלופיות אחדות להערכת ההוראה.

דירוג רטרופקטיבי, על ידי בוגרים של קורסים מסוימים, נמצא בקורלציה בינונית-עד-גבוהה עם תוצאות סקרי ההוראה שנערכו על ידי אותם סטודנטים כשלמדו בקורסים המקוריים. המסקנה של ממצא זה היא שהערכות הסטודנטים את המרצה והקורס אינם משתנים באופן משמעותי עם חלוף הזמן, בהערכה רטרופקטיבית. בכך מופרך אחד המיתוסים הנפוצים ביותר לגבי סקרי ההוראה - לעיתים קרובות מרצים שמקבלים דירוג נמוך טוענים שהסטודנטים לא מעריכים את הוראתם כיום אבל לאחר שיקבלו פרספקטיבה של הוראה/למידה בקורסים הבאים, ואפילו של התנסות במקום העבודה לאחר קבלת התואר, הם יעריכו הרבה יותר גבוה את ההוראה שקיבלו בקורס שבו דירגו נמו את המרצה.

(Ory & Ryan, 2001; Overall & Marsh, 1980; Theall & Franklin, 2001)

מחקרים אחדים מצאו קורלציות בינוניות בין סקרי ההוראה וכמה מהשיטות החלופיות האחרות כמו **הערכה ע"י עמיתים** (Feldman, 1989), **ממונים**, **מומחים ותצפיתנים אחרים**. ואולם אין עדויות מסיביות לתמיכה בכך לגבי כל אחת מהשיטות הללו.

באשר להערות כתובות, כאשר מספרן בקורס המוערך היה גדול מספיק (לפחות עשרות רבות) וסקר ההוראה בוצע בצורה ראויה, הן נמצאו בקורלציה גבוהה עם הדירוג המספרי.

V. סיבה נוספת להעדפת סקרי ההוראה על השיטות החלופיות – שמיעת קולם של הסטודנטים

גם מבחינה עקרונית, סקרי ההוראה עדיפים כשיטה בלעדית להערכת ההוראה על שאר השיטות החלופיות. זוהי השיטה היחידה מבין מבחר השיטות שמייצגת את דעותיהם של הסטודנטים לגבי ההוראה. קולם של הסטודנטים צריך וראוי להישמע כשמעריכים את המורים שלהם. מכלל בעלי העניין בהערכת ההוראה, הם היחידים שחשופים להוראה באופן מתמשך (לעומת תצפיות בשיעור למשל, שהן אירוע חד פעמי בדרך כלל) והם היחידים שמקיימים אינטראקציה עם המורה כמו השתתפות בשיעורים, הצגת שאלות למורה בכיתה ומחוצה לה, הגשת עבודות ומבחנים וקבלת תוצאותיהם ועוד. האינטראקציות הללו הינן חלק מרכזי מהתפקוד של המורה וחשוב לתת להן משקל בהערכתו. מרבית השיטות החלופיות כלל אינן לוקחות בחשבון רכיב זה או שמתחשבות בו בצורה בלתי משמעותית.

VI. השורה התחתונה לגבי שימוש בשיטות חלופיות לסקרי ההוראה

אם המטרה היא הערכת ההוראה לשם השוואה או לעשיית החלטות אישיות כמו קידום, אז עקב בעיות של היתכנות (עלות/תועלת) ושל תקפות ומהימנות, לא ניתן להשתמש באף שיטה חלופית להערכת ההוראה בקנה מידה משמעותי כשיטה בלעדית (standalone), אלא רק כתוספת לסקרי ההוראה ולתמיכה בהם. וכך, יש כיום קונסנזוס בין המומחים/חוקרים/בעלי עניין אחרים שסקרי ההוראה העשויים בצורה ראויה עולים לאין שיעור על כל השיטות החלופיות להערכת ההוראה האקדמית מבחינת היתכנות, תקפות ומהימנות, וייצוג הסטודנטים. עם זאת, השימוש בשיטה חלופית בנוסף לסקרי ההוראה לשם טריאנגולציה הוא מאוד מומלץ, במיוחד במקרים של עשיית החלטות לגבי אנשי סגל.

(Arreola, 2007; Benton & Cashin, 2012, 2014; Benton & Li, 2017; Benton & Ryalls, 2016; Berk, 2006, 2013; Hativa, 2014a, 2014b; McKeachie, 1997; Seldin, 2006; Theall, Abrami, & Mets, 2001).

לעומת זאת, אם מטרת ההערכה היא רק לתת למרצים הנחיות ורעיונות לשיפור ההוראה, ניתן להשתמש בכמה שיטות חלופיות במקום סקרי ההוראה.

מה בפוסט הבא?

אם לא תעלה לפני כן שאלה של אחד הקוראים, אציג בפוסט 10 מה ידוע מהמחקר על ההיבטים המרכזיים בהתנהגות המורה בשיעור שתורמים להוראה טובה.

רשימת קריאה

Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system*. (3rd ed.). San Francisco: Jossey-Bass.

Benton, S. L., & Cashin, W. E. (2014). Student ratings of instruction in college and university courses. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory & research* (Vol. 29, pp. 279-326). Dordrecht, The Netherlands: Springer.

Benton, S. L., & Ryalls, K. R. (2016). Challenging misconceptions about student ratings of instruction. *IDEA Paper*, 58.

Berk, R. A. (2006). *Thirteen strategies to measure college teaching: A consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Sterling, VA: Stylus.

- Berk, R. A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty administrators must know to protect themselves in employment decisions*. Sterling, Virginia: Stylus.
- Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, 30(2), 137-194.
- Hativa, N. (2014a). *Student ratings of instruction: A practical approach to designing, operating, and reporting, Second Edition*: Oron Publications.
- Hativa, N. (2014b). *Student ratings of instruction: Recognizing effective teaching, Second Edition*: Oron Publications.
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218-1225.
- Ory, J. C., and Ryan, K. (2001). How do student ratings measure up to a new validity framework? In M. Theall, P. C. Abrami, and L. A. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them? New directions for institutional research* (Vol. 109, pp. 27-44). San Francisco: Jossey-Bass.
- Overall, J. U., & March, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology*, 72, 321-325.
- Theall, M., Abrami, P. C., and Mets, L. A. (Eds.). (2001). *The student ratings debate: Are they valid? How can we best use them? New directions for institutional research* (Vol. 109). San Francisco: Jossey-Bass.
- Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? In M. Theall, P. C. Abrami, & L. A. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them? New directions for institutional research* (Vol. 109, pp. 45-56). San Francisco: Jossey-Bass.